AFRL-RI-RS-TR-2018-113

# CONVERSATION UNDERSTANDING THROUGH BELIEF INTERPRETATION SOCIOLINGUISTIC MODELING (CUBISM)

**THE FLORIDA INSTITUTE FOR HUMAN AND MACHINE COGNITION (IHMC)**

*MAY 2018*

FINAL TECHNICAL REPORT

***APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED***

STINFO COPY

**AIR FORCE RESEARCH LABORATORY**
**INFORMATION DIRECTORATE**

■ **AIR FORCE MATERIEL COMMAND** ■ **UNITED STATES AIR FORCE** ■ **ROME, NY 13441**

# NOTICE AND SIGNATURE PAGE



AFRL-RI-RS-TR-2018-113 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE CHIEF ENGINEER:

**/ S /**
JAMES M. NAGY
Work Unit Manager

**/ S /**
JON S. JONES
Technical Advisor, Information Intelligence Systems and Analysis Division
Information Directorate

# REPORT DOCUMENTATION PAGE

*Form Approved*
**OMB No. 0704-0188**



| **1. REPORT DATE** *(DD-MM-YYYY)* | **2. REPORT TYPE** | **3. DATES COVERED** *(From - To)* |
|---|---|---|
| MAY 2018 | FINAL TECHNICAL REPORT | SEP 2012 – NOV 1027 |

**4. TITLE AND SUBTITLE**

CONVERSATION UNDERSTANDING THROUGH BELIEF INTERPRETATION SOCIOLINGUISTIC MODELING (CUBISM)

**5a. CONTRACT NUMBER** FA8750-12-2-0348

**5b. GRANT NUMBER** N/A

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**

Wilks Yorick

**5d. PROJECT NUMBER** DEFT

**5e. TASK NUMBER** 12

**5f. WORK UNIT NUMBER** 06

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**

The Florida Institute for Human and Machine Cognition Inc.
15 SE Osceola Ave
Ocala FL 34471

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**

Air Force Research Laboratory/RIED
525 Brooks Road
Rome NY 13441-4505

**10. SPONSOR/MONITOR'S ACRONYM(S)** AFRL/RI

**11. SPONSOR/MONITOR'S REPORT NUMBER** AFRL-RI-RS-TR-2018-113

**12. DISTRIBUTION AVAILABILITY STATEMENT**

Approved for Public Release; Distribution Unlimited. This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT**

The CUBISM project aimed to bring together research related to dialogue understanding along two different analytical dimensions, namely (1) participants' mental content, and (2) participants' social roles and relationships. With respect to mental content, we originally intended to extend the ViewGen belief ascription system (Wilks and Ballim, 1990) to maintain beliefs and other propositional attitudes for individuals and groups and to model the change and exchange of beliefs/attitudes within and amongst individuals and groups. This work had already been extended in ONR work. The aim was to populate this "belief engine" with semantic content extracted from dialogues and attributed to participants, so as to track their changing beliefs as dialogues progressed. The plan was to be the first system in which a belief engine was combined with real content extracted from dialogue corpora. The belief engine was re-implemented but data availability became a problem but we were able to shift the emphasis of the project to knowledge basis population. The other major component by UAlbany, was successful in competitions on sentiment extraction. The work was based on the assumption that social relations such as 'leader' or 'influences' can provide important social information about participant roles and relationships. Such social information is latent in dialogue and derivable via sociolinguistic features as locator of sentiment. The major research achievement of the project has been to show that belief and sentiment change in dialogue are in fact correlated and one can be used as a predictor of the other.

**15. SUBJECT TERMS**

SENTIMENT, BELIEF, KNOWLEDGE-BASE-OPULATION, SOCIAL-INFLUENCE, DIALOGUE, ANOMALY

| **16. SECURITY CLASSIFICATION OF:** | | | **17. LIMITATION OF ABSTRACT** | **18. NUMBER OF PAGES** | **19a. NAME OF RESPONSIBLE PERSON** |
|---|---|---|---|---|---|
| **a. REPORT** | **b. ABSTRACT** | **c. THIS PAGE** | | | JAMES M. NAGY |
| U | U | U | UU | 25 | **19b. TELEPHONE NUMBER (*Include area code*)** |

## Table of Contents



## Table Of Figures

# 1 SUMMARY

The CUBISM project aimed to bring together research related to dialogue understanding along two different analytical dimensions, namely (1) participants' mental content, and (2) participants' social roles and relationships. ViewGen is a dynamic beliefs management system. It generates multiple belief environments from different points of view. The basic inference mechanism in ViewGen is default reasoning. That is, unless there is evidence to the contrary, the agents are assumed to have the same beliefs as the system. ViewGen allows multiple viewpoints. With respect to mental content, we originally intended to extend the ViewGen belief ascription system (Wilks and Ballim, 1990) to maintain beliefs and other propositional attitudes for individuals and groups and to model the change and exchange of beliefs/attitudes within and amongst individuals and groups. This work had already been extended in Office of Naval Research-supported work at the Florida Institute for Human and Machine Cognition (IHMC) (Wilks et al., 2014). Our aim was to populate this "belief engine" with semantic content extracted from dialogues and attributed to participants, so as to track their changing beliefs as dialogues progressed.

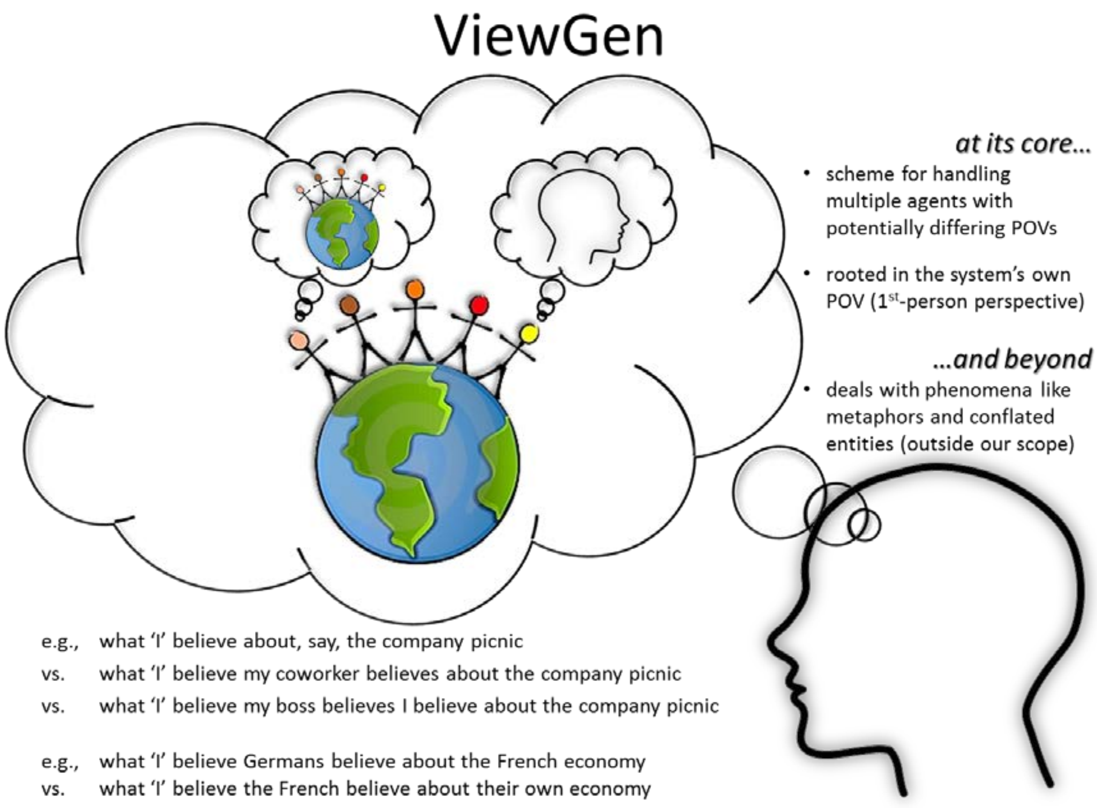


**Figure 1: VIEWGEN**

The other major component at the University of Albany (UA) was more independently successful in competitions on sentiment extraction: it was based on the assumption that social relations such as 'leader' or 'influences' provide important social information about participant roles and relationships. Such social information is latent in dialogue and derivable via sociolinguistic features such as topic, sentiment, and notions of "distance" between dialogue participants by our project partners (Shaikh et al., 2016). This strand of work has proved highly successful as a locator of sentiment and true dialogue "topic". The major research achievement of the project has been to show that belief change and sentiment change in dialogue are in fact correlated and one can be used as a predictor of the other. This is important since sentiment computation is much "cheaper" in computational terms than nested belief computation. These two forms of analysis (i.e., semantic and sociolinguistic) have normally been thought disjoint but we believed, correctly as we showed, that they can strengthen each other by bringing together the "internal" information about beliefs that is directly or indirectly expressed in utterance content with the "external" social information about believers that is signaled in overall dialogue behavior. Ultimately, we want a CUBISM system to answer questions like:

- What is the state of belief of dialogue participants?
- What is the nature of influence and leadership exerted in the dialogue?
- How quickly and in what directions are beliefs evolving?
- What topics are discussed and what are the attitudes of participants towards those topics?
- Are there anomalies indicative of, say intentional deception?

The third strand in our CUBISM work at the University of Florida (UF), was a learned optimization (Zhou et al., 2016) over large data of the fusion described above which we believe showed the scalability of the approach were it be presented and housed in an appropriate database.

In our hybrid approach to dialogue understanding, we sought to combine the effects of essentially symbolic/representational methods with empirical/data-driven methods to determine, for example, the likely sequence of topic shifts within conversations of a certain type, given the current social behavior and evolving positions of the participants. Our general hypothesis is that combining these methods will produce results that are more accurate and more robust to corrupted or missing data than either approach in isolation. The principal hypothesis of this work is that *social-computation (of sentiment, topic, role of participant etc.) can be used to identify areas in a dialogue, based on change of topic or role, where we will be able to focus the attention of the belief engine to seek corresponding changes in belief*. Since belief computation is intensive and potentially unbounded, changes in sociolinguistic features may mark out plausible target areas that constrain and scope the search space, just as, in classic Question Answering (QA), Information Retrieval (IR) computations identify passages of text where Information Extraction (IE) can then search for plausible answers.

# CUBISM Architecture

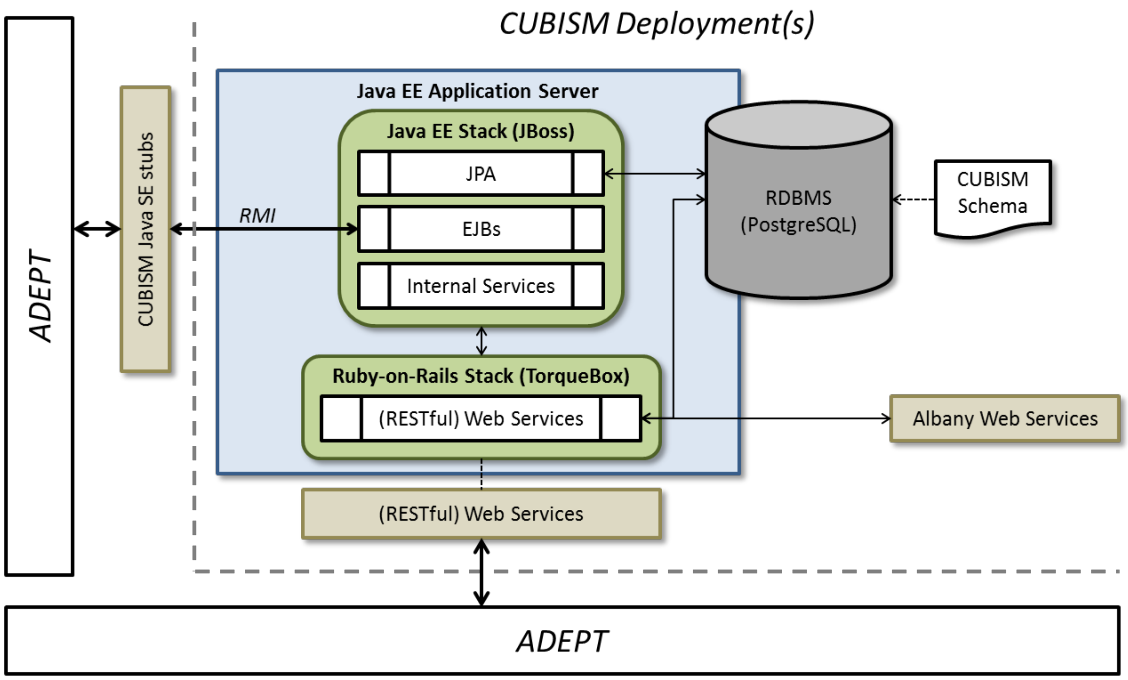


**Figure 2: CUBISM Architecture**

# 2 INTRODUCTION

Analysts of conversations (for their security implications) need to:

- determine the beliefs of the participants (what they assent to and deny etc.) as well as their dynamic changes during a conversation;
- determine the belief relations between participants, including quasi-logical relations of denial, reinforcement etc.;
- determine the content and function of a missing, unseen, interlocutor;
- determine the relative social roles of the conversational partners (leadership, acquiescence, etc.) and how these relationships change over time;
- determine the sentiment held by participants towards a target entity or event or relation;
- determine anomaly and possible deception between participants, based on models of topic flow in dialogue and expected causal and logical consequences.

We aimed for a conversational analysis system (including web based chatrooms and forums) that can do these things in English and other languages such as Spanish and Chinese, and provide information about dialogue content of obvious importance to analytic understanding of an interlocutor's (e.g., an opponent's) possible designs but until now attainable only through direct *human* attention to the conversations.

# 3 METHODS, ASSUMPTIONS AND PROCEDURES

## 3.1 Research Objectives

### 3.1.1 Problem Description

Analysts of conversations (for their security implications) need to:

- determine the beliefs of the participants (what they assent to and deny etc.) as well as their dynamic changes during a conversation;
- determine the belief relations between participants, including quasi-logical relations of denial, reinforcement etc.;
- determine the content and function of a missing, unseen, interlocutor;
- determine the relative social roles of the conversational partners (leadership, acquiescence, etc.) and how these relationships change over time;
- determine the sentiment held by participants towards a target entity or event or relation;
- determine anomaly and possible deception between participants, based on models of topic flow in dialogue and expected causal and logical consequences.

We aimed for a conversational analysis system (including web based chat rooms and forums) that can do these things in English and other languages such as Spanish and Chinese, and provide information about dialogue content of obvious importance to analytic understanding of an interlocutor's (e.g., an opponent's) possible designs but until now attainable only through direct *human* attention to the conversations.

### 3.1.2 Research Goals

The purpose of our algorithm for deep analysis was to enable the larger DEFT system to discover, summarize, and alert regarding meaningful dialogue content that is responsive to such queries as:

- What is the state of belief of dialogue participants?
- What is the nature of the influence and leadership exerted in the dialogue?
- How quickly and in what directions are belief states evolving?
- What is the sentiment of dialogue participants towards specific targets (entities, relations, and events) and how is it changing over time?
- What topics are discussed and what are the attitudes of participants towards those topics?
- What are the overall aggregated beliefs and sentiments of dialogue participants towards entities, events and relations?
- Are there anomalies indicative of intentional deception?

Our work feeds into a TA2 platform and quantificational issues will be addressed there.

While overall DEFT integration is the responsibility of the TA2 performer, some amount of coordinated development, integration, and standalone evaluation is necessary. We shall therefore create an integration platform at IHMC and mirrored at UA and UF, where the independent components of our algorithm are to be integrated and comparatively assessed. UF will contribute the platform's scalable, probabilistic knowledge-base and algorithms for integrating uncertain information from various sources and inferential models. We shall process substantial dialogue (and other) corpora both to derive social and linguistic models and derive background knowledge from non-dialogue corpora.

#### 3.1.3 Expected Impact

The project's goal was to enable the questions posed as research goals to be answered, and that will be a significant advance on the currently available technological support to analysts.

- Research grounded in Social Science theory
- Our objective: Modeling people, behaviors, and topics
  - This is not about individual documents or features
- Given multiple units of information:
  - Documents, postings, utterances, responses
- We want to find out:
  - Who are the key *participants*
  - What are their *roles* and mutual *relationships*
  - What are the *main topics* they discuss
  - What are their *positions* towards these topics
  - What do we know or can infer about their *beliefs*
- Are there anomalies involved in any of the above?

**Figure 3: CUBISM Approach**

### 3.2 Detailed Description of Technical Approach

The originalities of the approach are:

- Bringing together semantic content of dialogue utterances and a belief engine that computes the shifting beliefs of participants—actively populating a belief manipulation engine and knowledge base from real corpus data has not been done before.
- Bringing together semantic content of dialogue and social features of participants (influence, group membership, leadership, conviction etc.) computed over their non-linguistic features, is also original and these two forms of dialogue analysis (semantic and social) have not previously been effectively linked.

In operation, the proposed algorithm constructs an annotated database for conversations that indexes individual participants to the content of their beliefs, including possible shifts in belief during the course of conversation. The database will also contain pragmatic features such as:

- belief and sentiment relations between speakers and other entities, including reported belief and sentiment;
- the quasi-logical relationship of their beliefs to those of their interlocutors,
- anomalies (e.g., possibly deceptive features of specific utterances),
- sociolinguistic aspects of utterances, (e.g., attempted or successful acts of influence).

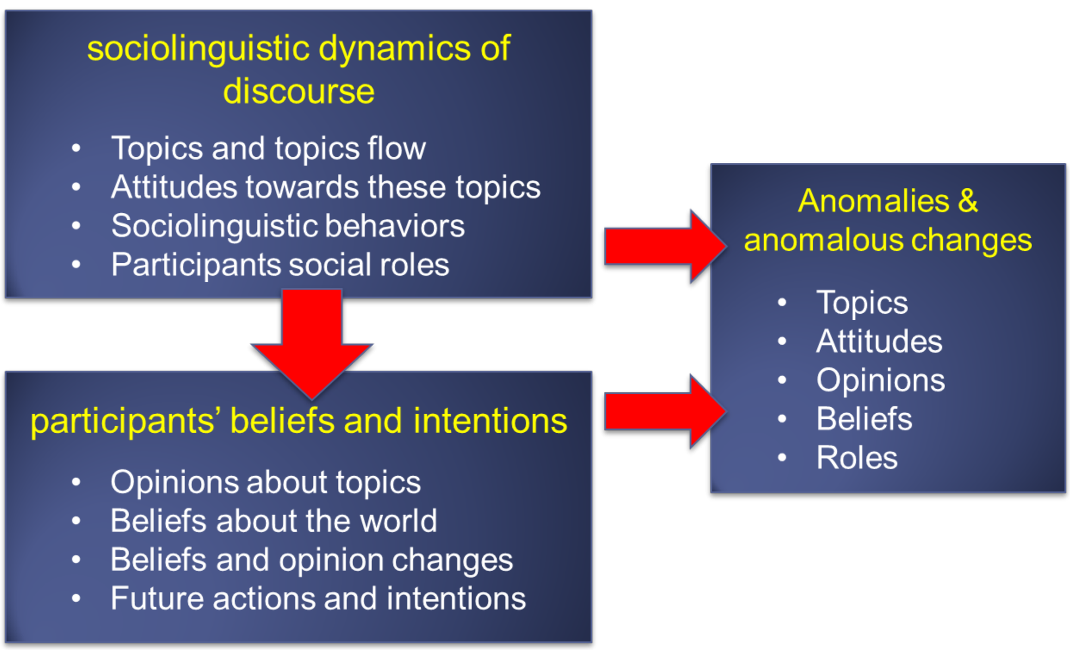


**Figure 4: CUBISM Components**

The effort leverages a number of existing technologies (some developed during prior collaboration of these same teams) to automatically spotlight key conversations and participants and to extract and summarize salient points in an annotated database. The heart of our approach is to interpret dialogue across a range of content levels — the first level being derived from an analysis of the surface text, and subsequent levels through interpretation by:

a. an existing ‘belief engine” that determines the relationships between the beliefs of interacting agents;
b. a new sentiment extraction system that determines the polarity and strength of sentiment from speakers and other agents towards other entities, as well as relations and events; and
c. a new belief extraction system that determines the strength and target of the beliefs held by speakers, including beliefs reported by others.
d. a reconfiguration of an existing conversational engine that applies empirically derived constraints to interpret utterances in terms of social function, such as position and role, within a whole discourse.

Meeting our goals as described above involved five phases of technical work:

Phase 1: Initial content extraction from transcribed conversations

Phase 2: Participants’ points-of-view belief and sentiment modeling

Phase 3: Conversation flow modeling and topic determination

Phase 4: Statistical determination of topic sequencing and constraints

Phase 5: Inter-participant social relationship determination

### 3.3 Populating the Knowledge Base

A key requirement of the Deep Exploration and Filter of Text (DEFT) program is population of the joint ADEPT Knowledge Base (KB). The CUBISM team algorithms populate content related to beliefs, sentiment, and social roles of individuals and other entities. We were instrumental in developing KB schema for population of sentiment and belief (the BBN ADEPT KB v2.2 is based on the design developed at Albany). We have implemented a working pipeline in the BBN ADEPT environment to insert sentiment and belief assertions in the KB. We have incorporated the UIUC Entity Co-reference system (ADEPT compliant Algorithm 10.8) (which is used to extract and populate entities) into our Target-Focused Sentiment Extraction algorithm (4.10) as well as an event/relation extraction system (also from UIUC or Stanford, not currently in ADEPT). We extract sentiment/belief assertions automatically from raw text on the entities, events and relations identified by the IE systems. In addition, our Sentiment algorithms (4.10) work on data that has been manually annotated for entities, events and relations to create and insert assertions to the KB. Algorithm 4.3 (Belief Modeling Algorithm) and 4.4 (Belief Ascription Algorithm) have both been implemented in the TAC approach, and have been incorporated in the pipeline and are inserting beliefs in the KB as of April 2016. Algorithm 4.15 (Consensus Maximization Fusion) our relation extraction fusion pipeline, and is currently undergoing the integration process, this algorithm incorporates relation extractors from other sites (such as the University of Illinois Urbana-Champaign (UIUC) and the University of Washington) and inserts a set of curated relations into the KB.

## 4 RESULTS AND DISCUSSION

In the course of the project, the CUBISM team made strong progress in research as well as in KB population readiness. On the research side, we developed what is currently the best sentiment system (based on TAC KB evaluations). The system was operational in English, Spanish and Chinese. We have made critical advances in belief extraction that identifies both belief source and target, in addition to the actual belief. We also pursued installing multiple beliefs/hypotheses into the KB, both statistically and in terms of our theory of belief environments (VIEWGEN). Our work was critical to establishing new joint Belief/Sentiment (BeSt) evaluation to be run early 2016 (LDC is currently preparing training data). We were also instrumental in developing KB schema for population of sentiment and belief (the BBN ADEPT KB v2.2 is based on the design developed at Albany). We worked with BeSt group to finalize KB population strategy. The team also participated in NIST evaluations in 2014 and 2015 and demonstrated top-tier performances. Most of our algorithms operate in all three program languages (English, Spanish, Chinese) and were inserted into ADEPT.

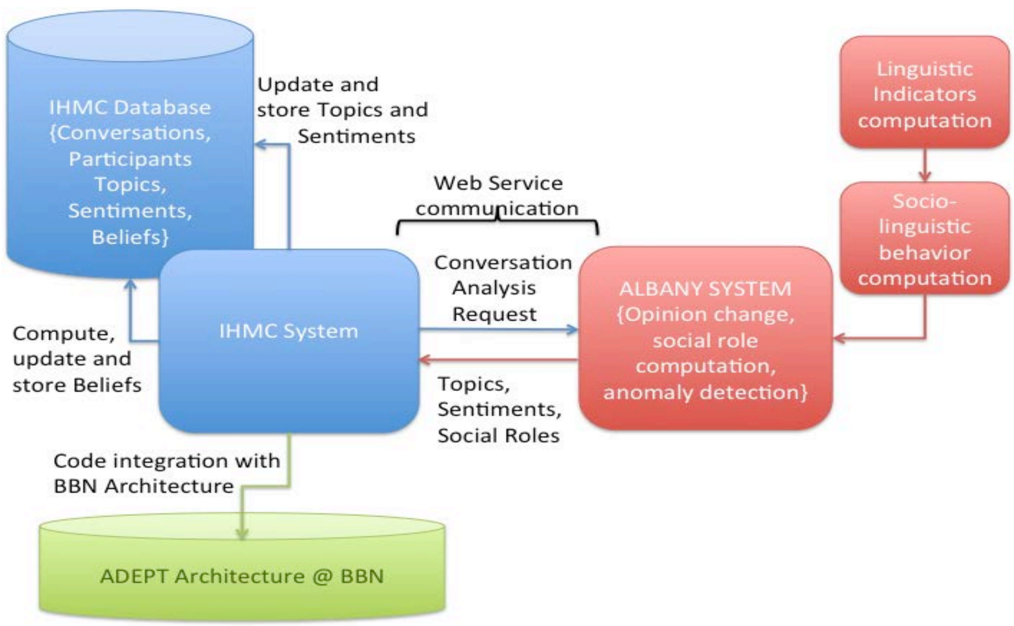


**Figure 5: ADEPT Implementation**

We implemented a working pipeline in the BBN ADEPT environment to insert sentiment and belief assertions in the KB, and incorporated the UIUC Entity Co-reference system (ADEPT compliant Algorithm 10.8) into our workflow (which is used to extract and populate entities) as well as an event/relation extraction system (also from UIUC or Stanford University, not currently in ADEPT). We extracted sentiment/belief assertions automatically from raw text on the entities, events and relations identified by the IE systems. These are algorithms 4.1 and 4.11 in the ADEPT API. In addition, our Sentiment and Belief algorithms work on data that has been manually annotated for entities, events and relations to create and insert assertions to the KB. Algorithm 4.3 (Belief Modeling Algorithm) and 4.4 (Belief Ascription Algorithm) have both been implemented in the TAC approach. In addition, our ensemble algorithm was validated in the TAC SFV evaluation. Our 3 submitted runs achieved top 3 overall performance in the 2015 KBP evaluation among over 69 runs from over 20 teams and other SFV filter/ensemble teams.

The following algorithms were operational on BBN ADEPT interface

1. Topical Positioning (4.6)
2. Target Focused Sentiment Extraction (4.1)
3. Target Focused Belief Extraction (4.11)
4. Influencer Finder (4.12)
5. Leader Finder (4.13)
6. PowerPursuit Finder (4.14)
7. Belief Modeling (4.3)
8. Belief Ascription (propagation) (4.4)
9. Multiple probabilistic assertion fusion of sentiment/belief from different sources.

## 4.1 DEFT Evaluations

The team participated in NIST/TAC evaluations including KBP-Sent and TAC/SFV. Our sentiment system took the #1 spot in 2014, and our 3 submitted SFV runs achieved top 3 spots in the 2015 KBP evaluation among over 69 runs from over 20 teams and other Slot Filler Evaluation (SFV) filter/ensemble teams.

The CUBISM team formed part of the BeST group working on the combined Belief/Sentiment evaluation in 2016. The evaluation guidelines were designed by the team, and a pilot evaluation was run in late 2014.

## 4.2 Comparison with Current Technology

What we proposed was innovative and original with respect to the state of the art:

- The initial IE phase of content extraction from dialogue (though this may prove dispensable in the second-phase, TA2 provided platform).
- The determination of dialogue structure by means of a virtual dialogue machine based on stacked networks — though a similar system was deployed by Lemon et al. in the WITAS project at Stanford.
- The ViewGen belief engine for modeling the beliefs agents hold about each other in dialogue — though there have been other systems, since the original work of Allen and of Perrault, using a space-based metaphor and agent models for computing belief.
- The Sentiment Extraction system developed for DEFT is currently the top performing (by far) system as evidenced by NIST 2014 evaluations. It is currently the only sentiment system that can effectively extract targeted sentiment and embedded sentiment, a significant advance over traditional lexical systems. In addition, our system operates in several languages including English, Spanish, and Chinese.
- The work on topic tracking and social relationships in dialogue at UA that exposes sociolinguistic behaviors and group dynamics among the dialogue participants, including task and topic control, social positioning, leadership, and influence.
- The Slot Filling Validation system based on new probabilistic fusion and ensemble models developed for DEFT is currently the top performing – our 3 runs took the top-3 overall performance in the 2015 KBP evaluation among over 69 runs from over 20 teams and other TAC SFV filter/ensemble teams.

- The work on probabilistic knowledge-base construction and large-scale statistical inference at UF that implements statistical text analysis algorithms (e.g., extraction, inference) over large volumes of text data and provides principled means for storage, query, and analysis of uncertain data.
- The work on ensemble over knowledge extractions at UF proposed and developed an SFV system that uses a semi-supervised ensemble learning approach to aggregate the results from multiple slot fillers or text extractors.
- Work on the detection of anomaly in corpora by statistical methods developed by Guthrie and Wilks under UK Ministry of Defence sponsorship.

We believe this combination of techniques, largely originated by the proposers, yielded an algorithm capable of determining rich and complex relationships in conversation and reveal their underlying structure to analysts when integrated into the TA2 DEFT system.

Below, technical accomplishments are described.

*Event Argument Extraction*

We presented our event detection system, DISCERN, at TAC KBP 2015. Two contributions of this work are: (a) a web interface that improved efficiency during system development by enabling quick changes to linguistic rules and examination of their effect on precision and recall at runtime and (b) an approach to collapsing support-verb and event nominals that improved recall of argument linking. Combining linguistic knowledge with machine learning methods is shown to improve performance over either method alone. The definition of belief we are using means that a substantial subset of relevant beliefs can be detected using a modification of this event-seeking strategy. Beliefs will be injected into the ADEPT knowledge base as a relation between events and their argument fillers (found through the DISCERN system) and the entity believers (assigned through our belief ascription algorithm).

*Nested Reported Beliefs*

Previous incarnations of our belief extraction algorithm only dealt with first-person beliefs, i.e. simple beliefs about the world held by the speaker. We have extended the belief extraction algorithm to deal with reported beliefs (both second- and third-person reports). Thus, the algorithm now handles nested beliefs such as, “Mary thinks that the shirt is over-priced.” For this example, belief extraction extracts that the speaker believes that Mary believes that the shirt is over-priced. In general, the extraction algorithm decomposes belief-bearing statements into a sequence of believers (e.g., [speaker *believes*, Mary *believes*]) and the propositional content of the reported belief (e.g., ‘the shirt is over-priced’). The extracted sequence of believers is used to attribute the reported belief the appropriate nested point of view (in this case, the speaker’s point-of-view of Mary’s point-of-view).

*Propositional Attitudes besides Belief*

Earlier versions of CUBISM dealt only with beliefs, as opposed to other propositional attitudes such as intentions or desires. We have extended our belief extraction algorithm and our ViewGen-based representation scheme to handle multiple modalities of propositional attitudes. With respect to extraction, the process is very similar to belief extraction except that we incorporate a catalog of modal verbs with which to detect propositional attitudes beyond simply 'belief'; and extraction produces a sequence of agent-modality pairs (representing nesting scopes) and the non-modal propositional content. With respect to representation, agent point-of-views are now explicitly associated with a modality, while before the modality was tacitly assumed to be 'belief'.

*Target-Focused Sentiment Extraction*

The CUBISM team has developed sentiment extraction system that is currently the best state-of-art based on 2014 TAC KBP Sentiment evaluation. The system employs a novel approach to capture the sentiment of an entity towards another entity. We capture the contributing elements of sentiment, namely the sentiment holder, the sentiment target and the sentiment relation in an Affect Calculus. The semantic roles of the sentiment source, target and the relation are used to determine that sentiment in the Affect Calculus.

*Meso-topics as an Integration Point for Socio-linguistics and Belief Modeling*

Socio-linguistic analysis and belief models now have a shared concept of topic. As in prior versions of CUBISM, our socio-linguistic algorithms automatically identify the main topics in a conversation and annotate each statement with the main topics mentioned therein. These topics, and participants' attitudes toward them, are key measurements for our socio-linguistic analysis. We have now integrated these topics into our belief models. In particular, extracted beliefs are organized within an agent's point of view according to the topic(s) of the source utterance. This shared concept of topic will facilitate cross analysis of belief and socio-linguistic features (since, e.g., semantic content such as beliefs about a topic can be cross-indexed with socio-linguistic measures such as sentiment toward a topic, with both of modeled for each participant).

*Topical Positioning and subgroups*

We refined our algorithms for computing Topical Positioning measures and its component indices. Using human annotated data as validation, we adjusted thresholds for checking valence of words in valence lexicons. We also developed a method of tracking participants' positions over time on meso-topics and contrasting these with the positions of other participants. By plotting positions on certain meso-topics for select participants, we can track potential changes of belief for participants; we have implemented algorithms to compute these from (automatically) annotated dialogues marked for topic and polarity. We have developed a preliminary approach to identify subgroups in conversations based on their positioning on the main topics. Several NIST dialogues were annotated for subgroups, including extreme and moderate positions (for or against).

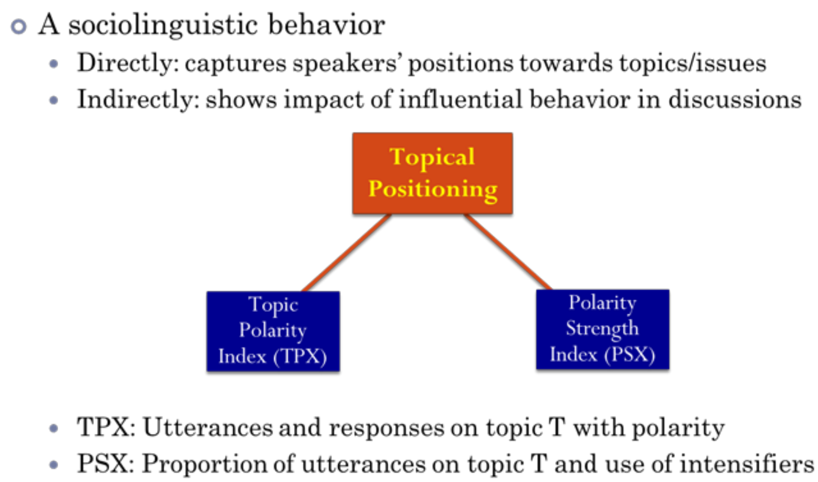


**Figure 6: Topical Positioning Measure**

In addition, work continued on refining the sentiment extraction algorithm. Specifically, we replaced a number of temporary heuristics with more finalized methods for target identification as well as for returning the correct offsets that has depressed our results in last year evaluation. Based on training data, our performance has increased to approximately 56% F-score (vs. 25% last year).

*Ensemble Knowledge Extraction*

The proposed SFV system uses a semi-supervised ensemble learning approach to aggregate the results from multiple slot fillers from the Cold Start track. We apply Consensus Maximization (CM) to combine the output of supervised stacked ensemble methods with the output of slot filling runs that can't be trained. By using CM we are also able to leverage a small set of assessed fillers to increase the performance of the system. The ensemble results outperformed the best cold start run, the best filtered runs, and other ensemble systems.

Current approaches on multi-modal extraction mostly focus on developing models to fuse low-level features from multiple modalities and learn unified representation from different modalities. But most related work failed to justify why we should use multimodal data and multimodal fusion, and few of them leveraged the complementary relation among different modalities. We identify the correlative and complementary relations among multiple modalities and propose a probabilistic ensemble fusion model to capture their complementary relation (images and text). Experimental results on the UIUC-ISD dataset show our approach outperforms approaches using only single modality. Word Sense Disambiguation (WSD) is the use case we studied to demonstrate the effectiveness of our probabilistic ensemble fusion model.

*Large-Scale Probabilistic Knowledge Base Learning and Inference*

Due to the limitations of human knowledge and extraction algorithms, current knowledge bases are still far from complete. We studied the problem of mining first-order inference rules to facilitate knowledge expansion. We proposed the Ontological Pathfinding algorithm (OP) that scales to web-scale knowledge bases via a series of parallelization and optimization techniques, including a new parallel rule mining algorithm implemented on Spark, a novel partitioning algorithm to break the learning tasks into smaller independent subtasks, and a pruning strategy to eliminate unsound and resource consuming rules before applying them. Combining these techniques, we were able to develop a first rule learning system that scales to Freebase. We mine 36,625 inference rules in 34 hours; no existing system achieves this scale.

Further, we developed General Iterative State Transition (GIST), a new database operator for parallel iterative state transitions over large states. GIST receives a state constructed by a UDA, and then performs rounds of transitions on the state until it converges. A final UDA performs post-processing and result extraction. We argue that the combination of UDA and GIST (UDA-GIST) unifies data-parallel and state-parallel processing in a single system, thus significantly extending the analytical capabilities of DBMSes. We exemplify the framework through two high-profile applications: cross-document co-reference and image de-noising. We show that the in-database framework allows us to tackle a 27 times larger problem than solved by the state-of-the-art for the first application and achieves 43 times speedup over the state-of-the-art for the second application.

### 4.3 Sentiment work during 2017 at the University of Albany

During 2017 funding had ceased for work at the IHMC and UF sites but the UA team was able to continue until the end of the project and to optimize their work on determining sentiment between a target (the sentiment “holder”) and a target, about which the sentiment was felt. Sentiment is observable in language through interaction dynamics and semantic role modeling, where it manifests as feelings, attitudes, emotions or opinions. These manifestations capture the subjective impressions, which usually assumes a binary opposition in opinions such as good/bad, like/dislike, for/against, etc. Our goal was to determine the attitude a speaker has towards another person, an object, an event or other appropriate entities. Being able to detect and extract sentiment from a discussion enables us to predict and determine the participant’s behavior. Additionally, it enables us to conduct an in-depth study of idea propagation through groups, since opinions and reactions to ideas are relevant to adoption of new ideas. Analyzing sentiment reactions on blogs can give insight to this process. It also poses the potential to augment psychological investigations/experiments with data extracted from natural language text, such as dream sentiment analysis. In general, humans are subjective creatures and opinions are important.

During this evaluation, we used the 2016 Sentiment training and test data as our current training data, assuming that they can be treated equally. We focused on English documents, and all data sets were provided by the Linguistic Data Consortium 1 (LDC), annotated with entities, relations and events (ERE). The ground truth, sentiment polarity among annotated pairs, was provided for the training data. Our approach automatically extracted sentiment from natural language text, along with the sentiment holder and the sentiment target.

### 4.3.1 Developments for Chinese Dialogues

We used the Stanford Chinese Parser and Sentiment Corpus (HowNet knowledge Database[1]) to generate Topic Polarity Index (TPX), Polarity Strength Index (PSX) and Topical Positioning Measure (TPM) measures for each participant. We then took all TPMs to form the Topical Positioning Network (TPN) where each node represents a participant and edge is the topical distance between two TPMs computed by cosine measure. We applied TPN to predict opinion changes over the SCIL Chinese chat data. The system is run to compute influencer score to derive the influencer for a dialogue. We then calculated before and after topical distance, cosine measure, between influencer and each participant denoted by $COSINE_{before}$ (TPM(Influencer), TPM(X)) and $COSINE_{after}$ (TPM(Influencer), TPM(X)). If ($COSINE_{before}$ - $COSINE_{after}$) is negative, it means participant X moved closer to the Influencer. If the value is positive, participant X moved away from influencer. We are primarily interested in those participants drastically moving away from influencer. We validated the result with human annotators.

### 4.3.2 Developments for Spanish Dialogues

We implemented preliminary capabilities for processing Spanish dialogues. We are using Spanish POS tagger and Spanish ANEW+ lexicon to compute Topical Positioning measures for participants in Spanish dialogues. The Spanish modules in system currently operate in alpha mode, where some capabilities such as co-reference resolution and dialogue act tagging are not fully functional. We are testing the system on a set of Spanish dialogues that were carefully selected by native Spanish speakers and annotated for sociolinguistic phenomena such as influence, topical positioning and subgroup presence.

### 4.3.3 Developments for KB Population and ADEPT Integration

We developed the KB schema for population of sentiment and belief (the BBN ADEPT KB v2.2 is based on the design developed at Albany). We have worked with BeSt group to finalize KB population strategy. We continue working with BBN and other DEFT contractors (entity and event KB population teams) to make the KB population process optimal. Most of our algorithms operate in all three program languages (English, Spanish, Chinese) and are inserted into ADEPT or undergoing final benchmark tests.

We have also created a software pipeline that works entirely in the ADEPT environment to create and insert assertions regarding sentiment and belief into the DEFT KB from raw text.

## 4.4 Significant Changes in our technical approach in the course of the project

1. In our original technical approach, we planned to use the GATE (General Architecture for Text Engineering) NLP toolkit. Considering increased commonality with other DEFT TA1 performers, we moved to the use of the Stanford NLP toolkit instead of GATE.
2. In our original technical approach, we planned for parallel development in Phase 1 of a DAF-based dialog manager alongside the system's belief and socio-linguistic modeling capabilities. Based on the DEFT program's later goals and growing emphasis on anomaly, we deferred development of the dialog manager until a later program phase so as to focus on capabilities that are best aligned with later DEFT goals. We were never able to return to the dialogue work in the funded period.
3. We undertook new work found to be essential but not in the original work plan on the automatic derivation of stereotypical belief sets for classes of dialogue participants---- this was essential for the operation of the belief ascription algorithm, and led by Greg Dubbin (IHMC). We derived these classes by clustering participants according to shared beliefs. The belief ascription algorithm ascribes likely beliefs to participants according to the derived classes. These stereotypical beliefs reflect unstated propositions crucial to reasoning over beliefs.
4. From May 2014 to December 2014, we completed a seed effort on "Disambiguate Knowledge Extraction Using Multimodal Data" led by Daisy Zhe Wang (University of Florida). The seed effort was a feasibility study to use multimodal data related to the same event or entity to improve the quality of information extraction using an Extract-Link-Integrate process. As a result of the seed effort, we collected large multimodality data from the traditional and social media for the 2014 Brazil World Cup event, and we identified four use cases and possible solutions for using multimodality data, including text, images, existing knowledge bases (KBs) such as WordNet and video, to improve the quality of information extraction results in terms of correctness and completeness.
5. From Jan 2015 to August 2015, as a continuation of the 2014 VITA seed effort, we completed a second seed effort on "Let there be confidence: Probabilistic Fusion of Multimodality Extractions" led by Daisy Zhe Wang (University of Florida) to explore the possibility of extending the multimodal extraction and integration models with confidence values. We explored probabilistic fusion models over multimodal extraction to further showcase the potential benefit of multimodality data sources and results from various analytics in improving the quality of the extracted information in terms of trustworthiness and confidence. We also showed uses to use multimodal data in discovering contradictions in data sources

## 5 CONCLUSIONS

Our plan was to be the first system in which a belief engine was combined with real content extracted from dialogue corpora. As always achievement fell short of intention: the belief engine was re-implemented, but data availability became a problem from the very outset of the project as it so often does. In addition, the sponsors designed a platform and data formats that made it very hard to represent the kind of structured information that this approach required. However, we were able to conform to that and to shift the emphasis of the project to knowledge basis population as we were required to do. At every stage, we made the required deliverables. However, much of the original impetus and original approach of the project thus became lost, since it was to fuse the approaches of three research traditions (at IHMC UAlbany and UFlorida) to produce a unique fused output that would track and predict the dialogue and mental movements of groups. Within that project data base format-----designed for quite different more conventional approaches—that became almost impossible and we had to pursue fusion within our own data structures.

## 6 REFERENCES


Bhatia, A., Dalton, A., Dorr, B., Dubbin, G., Hollingshead, K., Kandaswamy, S., and Perera, I. "Event argument linking and event nugget detection task: IHMC DISCERN system report" In *Proceedings of NIST TAC KBP*. 2015

Chen, Y., and Wang, D. "Knowledge Expansion over Probabilistic Knowledge Bases", *Proceedings of the ACM SIGMOD International Conference on Management of Data* (pp. 649–660). New York, NY: ACM Press. 2015

Chen, Y., Goldberg, S., Wang, D. and Johri, S. "Ontological Path Finding: Mining First-Order Knowledge from Large Knowledge Bases" *Proceedings of the ACM SIGMOD International Conference on Management of Data, 2016.*

Chen, Y., Goldberg, S., and Wang, D "ScaLeKB: scalable learning and inference over large knowledge bases" *The VLDB Journal*: 1-26. 2016

Chen, Y., Goldberg, S., and Wang, D "Archimedes: Efficient Query Processing over Probabilistic Knowledge Bases" *ACM SIGMOD Record*46.2 30-35. 2017.

Clark, M., Dalton, A., By, T., Wilks, Y., Shaikh, S., Lin, C., & Strzalkowski, T "Influence and Belief in Congressional Hearings", Participation paper for the 2014 NLP Unshared Task in *PoliInformatics*. 2014

Dalton, A., Wixted, M., Wilks, Y., Alagesan, M., Katsios, G., Subburathinam, A., and Strzalkowski. T. "Target-focused sentiment and belief extraction and classification using CUBISM" *Proceedings TAC-16, Washington DC.* 2016.

Dubbin, G., Bhatia, A., Dorr, B. Dalton, A., Hollingshead, K>, Perera, I., Kandaswamy, I.,and Hwang, J. "Event Nugget Detection and Argument Extraction with DISCERN" In *The Twenty-Ninth International Flairs Conference*. 2016.

Gong, D, Wang, D., and Peng, Y. "Multimodal Learning for Web Information Extraction" *ACM International Conference on Multimedia*. 2017.

Gong, D. and Wang, D. "Extracting Visual Knowledge from the Web with Multimodal Learning", IJCAI. 2017.

Katsios, G. A., Alagesan, M., Sharma, A. and Strzalkowski, T. "Target Focused Sentiment Extraction Framework" *Proceedings of TAC 2017 Workshop, National Institute of Standards and Technology*, Gaithersburg, MD, Nov. 13-14, 2017.

Liu, T. Cho, K., Strzalkowski, T., Shaikh S., and Mirzaei, M. "Validation of MRCPD Cross-language Expansions on Imageability Ratings". *LREC-2016 Conference, Slovenia.*2016

Peng, Y., Wang, D., Patwa, I., Gong, D., and Fang, V. "Probabilistic Ensemble Fusion of Multimodal Word Sense Disambiguation" *Proc. IEEE International symposium on multimedia*, 2015.

Peng, Y., Wang, D., Patwa, I., Gong, D., and Fang, V. "Scalable Image Retrieval with Multimodal Fusion" *Proceedings of the 29th International FLAIRS conference,* 2016

Peng, Y., Wang, D., Patwa, I., Gong, D., and Fang, V. "Probabilistic Ensemble Fusion of Multimodal Word Sense Disambiguation." *Proc. IEEE International symposium on multimedia*, 2015.

Ravishankar, V, Shaikh, S., and Strzalkowski. T. "Determining Political Affiliations from Conversational Data using Topical Positioning." Proc. *2016 International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction and Behavior Representation in Modeling and Simulation (SBP-BRIMS) UCDC Center Washington DC.*201

Ravishankar, V, Shaikh, S., and Strzalkowski. T. "Determining Political Affiliations from Conversational Data using Topical Positioning." Proc. *International Conference on Behavioral, Economic, and Socio-Cultural Computing (BESC 2016).* 2016

Rodriguez, M., Goldberg, S., and Wang, Z. "Consensus Maximization Fusion of Probabilistic Information Extractors." *The Proc. 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL 2016).* 2016

Rodriguez, M., Goldberg, S., and Wang, Z. "SigmaKB: multiple probabilistic knowledge base fusion." *Proceedings of the VLDB Endowment* 9.13 (2016): 1577-1580. 2016

Shaikh , S., Giarrusso, R., Ravishankar, V., and Strzalkowski, T." The SUNY Albany Sentiment Slot Filling System." In *Proceedings of the 2014 TAC KBP Workshop*. *NIST*. Gaithersburg, Maryland. 2014

Shaikh, S., Strzalkowski, S., Giarrusso, J., & Ravishankar, V. (2014). SUNY-Albany Sentiment Extraction System at the TAC 2014 Sentiment Slot Filling Evaluation Track. To appear in *Proceedings of the 2014 Text Analysis Conference*. Gaithersburg, MD: NIST Press. 2014

Shaikh, S. Cho, K., Strzalkowski, T., Feldman, L., Lien, J., Liu, T. and G. Broadwell "ANEW+: Automatic Expansion and Validation of Affective Norms of Words Lexicons in Multiple Languages" *LREC-2016 Conference, Slovenia*. 2016

Shaikh. S., Ravishankar, V., and Strzalkowski, T. "Towards a Framework for Determining Influence across Multiple Conversations" *Proc. 26th International Conference on Computational Linguistics (COLING 2016).* 2016

Wilks, Y., Clark, M., By, T., Dalton, D. and I. Perera "CUBISM: Belief, anomaly and social constructs In Marjorie McShane (ed), *Mental Model Ascription by Intellgent Agents*" Special issue of *Interaction Studies* 15:3 pp. 388 – 403, 2014.

Zhou, X., Chen, Y., and Wang. D. "ArchimedesOne: query processing over probabilistic knowledge bases" *Proceedings of the VLDB Endowment* 9.13 (2016): 1461-1464. 2016

## 7 ADDITIONAL REFERENCES

Wilks, Y. and Ballim, A. (1990) Relevant Beliefs. *In Proc. of European Conference on Artificial Intelligence*, Stockholm.

## 8 LIST OF ACRONYMS

**ADEPT: A DEFT** platform

**BBN:** Bolt, Beranek and Newman

**BEST**: Belief and Sentiment

**CM:** Consensus maximization

**CUBISM:** Conversational Understanding through Belief Interpretation and Sociolinguistic Modelling

**DBMS:** Data Base Management System

**DISCERN:** Discovering and Characterizing Emerging Events

**GIST:** General Iterative State Transition

**IE:** Information Extraction

**IHMC:** Florida Institute of Human and Machine Cognition

**IR:** Information Retrieval

**KB:** Knowledge Base

**LDC:** Linguistic Data Consortium

**NIST:** National Institute of Standards

**NLP:** Natural Language Processing

**ONR:** Office of Naval Research

**OP:** Ontological Pathfinding algorithm

**POS:** Part of Speech

**PSX:** Polarity Strength Index

**QA**: Question Answering

**SFV:** Slot Filler validation method

**TAC:** Text Analysis Conference

**TPN:** Topical Positioning Network

**TPM:** Topical Positioning Measure

**TPX:** Topic Polarity Index

**UIUC**: University of Illinois

**UDA**: Usable Data Abstraction

**VIEWGEN**: IHMC belief modeling platform.

**WSD:** Word Sense Disambiguation